Benchmarking of the 3DRISM and COSMO tools implemented in ADF package

Volodymyr Sergiievskyi

October 31, 2012

Abstract

In this study we investigate the accuracy of solvation free energy (SFE) calculations with the 3DRISM-SCF and COSMO methods implemented in the ADF package. We compare the calculated results to the experimental ones and to the results of 3DRISM standalone calculations performed in RISM-MOL-3D program package. We report the abnormally high absolute values of MDC-q partial charges (standard charges for ADF) for at least 23 of 99 organic compounds. We show, that high absolute values of the MDC-q charges correlate to the abnormally low values of SFE calculates with the 3DRISM-SCF method in ADF.

We use semi-empirical 3DRISM solvation free energy corrections performed with the Universal Correction (UC) model for all the molecules with reasonable partial charges. We show that the free coefficient in the UC model strongly depends on the choice of the training set and does not improve accuracy of SFE calculations. We propose to use the UC_0 with only one partial molar volume coefficient. We show, that 3DRISM SFE calculations with the UC_0 model with CHELPG or MDC-d partial charges predict the solvation free energy with accuracy of about 1 kcal/mol, which is better than accuracy of COSMO calculations.

1 Free Energy Calculations in ADF

ADF is a package for QM calculations which provides a possibility to model solvated molecules. Several methods for modeling solvation are implemented in ADF. In this work two of them: COSMO and 3DRISM are compared and tested on a set of organic molecules. In the COSMO solvation model the solute molecule is embedded

into the cavity in dielectric with a given electrical permutivity. A charge distributions, electrostatic potential and solvation free energy (SFE) can be calculated using the Poisson Boltzmann equation. In the 3DRISM solvation model implemented in ADF the solvent is represented in the RISM approximation. Electron densities are calculated self-consistently, accounting the solvent response to the solute. In this work we refer this procedure as *3DRISM-SCF* method. Typical calculation time for COSMO calculations is 1-2 minutes. 3DRISM-SCF procedure requires about 50 iteration steps to converge. Each step is a single-point 3DRISM calculation, which takes 1-5 minutes. Typical time of calculations for 3DRISM-SCF procedure is about 1 hour per molecule. All 3DRISM-SCF and COSMO calculations in this work were performed with the ADF package. To stress this we refer these methods in our work as COSMO_{ADF}, 3DRISM-SCF_{ADF}.

Although 3DRISM-SCF procedure is more rigorous from the theoretical point of view, it takes much more time than COSMO_{ADF} calculations. The way to reduce the computational expenses is to use *standalone 3DRISM* calculations. In this case the 3DRISM equations are solved only once for the given partial charges. This method does not account the changes of the partial charges due to the presence of the solvent. However, this method is much faster than time-consuming 3DRISM-SCF procedures. All standalone 3DRISM calculations in this work were performed with the multi-grid 3DRISM solver implemented in the RISM-MOL-3D package [1, 2]. To stress this in this work we refer the standalone calculations as $\text{3DRISM}_{RISM-MOL}$.

The 3DRISM theory provides the end-point expressions for the SFE calculations. However, it is known, that these expressions have several systematic errors which need to be corrected. The way to correct the errors provides the Universal Correction (UC) model. In the UC model the partial molar volume correction is used, which allows one to calculate the SFE with the accuracy of 1 about kcal/mol [3]. In our study we perform the 3DRISM-SCF_{ADF} and COSMO_{ADF} and standalone $\text{3DRISM}_{RISM-MOL}$ calculations. We use the UC model to correct the SFEs calculated with the 3DRISM methods. We compare the calculated SFEs to the experimental results and discuss the performance of different methods.

2 Methods

To test the algorithms the set of 99 organic compounds from Ref. [1] was used. For these compounds the experimental free energies are known. The full list of molecules with the experimentally measured SFEs one can find in Table 6.

Both: 3DRISM-SCF_{ADF} and standalone $\text{3DRISM}_{RISM-MOL}$ calculations require LJ parameters to be assigned to the atoms of the solute molecules. The molecules

Atom	σ [Å]	ϵ [kcal/mol]
H	2.5	0.03
C	3.5	0.066
O	2.9	0.14
Cl	3.4	0.3

Table 1: Lennard Jones parameters for atoms of different types

in the set contain Hydrogen, Carbon, Oxygen and Chlorine atoms. We used the same LJ parameters for all atoms of the same type (see Table 1). The standalone 3DRISM$_{RISM-MOL}$ calculations also require the partial charges to be assigned to the atoms of the solute molecule. The charges for standalone 3DRISM$_{RISM-MOL}$ calculations are typically determined from the QM calculations. ADF package allows one to calculate Multipole Derived Charges (MDC) [4]. Three versions of the MDC charges can be calculated with the ADF package: MDC-m, MDC-d and MDC-q charges. It is known, that RISM and 3DRISM calculations are sensitive to the choice of partial charges[5, 6, 7]. It is also known, that RISM calculations with CHELPG charges (CHarges Derived from Electrostatic Potential, Grid method) are good for solvation free energy calculations. Unfortunately, CHELPG charges cannot be calculated with ADF package. In the current work the ORCA Quantum mechanical software was used for calculation of the CHELPG charges[8].

The standalone 3DRISM$_{RISM-MOL}$ calculations were performed with RISM-MOL-3D package [2]. The calculations were performed for all the molecules in the set for four different partial charges (MDC-m,MDC-d,MDC-q,CHELPG). The Universal Correction (UC) model was used for calculation of the SFEs[3]. The UC model includes two corrections: Partial Molar Volume (PMV) correction, and free coefficient. The formula to calculate the SFE with the UC model in the following:

$$\Delta G_{UC} = \Delta G_{3DRISM} + a_1 \rho V + a_2 \tag{1}$$

where ΔG_{3DRISM} is uncorrected SFE calculated with one of 3DRISM expressions, V is PMV, $\rho = 0.0336 \mathring{A}^{-3}$ is the number density of water, the coefficients a_1 and a_2 are determined using the least squares method on the training set of compounds. Partial molar volume can be calculated within the 3DRISM framework.

3 Technical details

CHELPG charges were calculated using the B3LYP method and standard parameters in ORCA.

ADF COSMO calculations were performed for the standard parameters for water, i.e. dielectrical permutivity 78.4, SES (Solvent-Excluding-Surface) and the radius of the solvent molecule 1.93 Å.

3DRISM-SCF$_{ADF}$ calculations were performed for MSPC/E water with σ_H=1.0Å, ϵ_H=0.046 kcal/mol on the grid of size $64 \times 64 \times 64$ points with the grid step 0.5Å.

Standalone 3DRISM$_{RISM-MOL}$ calculations were performed for the MSPC/E water with σ_H=0.8Å, ϵ_H=0.046 kcal/mol on the grid with buffer 15Åand grid size 0.5Å.

In the both: 3DRISM-SCF and standalone 3DRISM$_{RISM-MOL}$ calculations the KH closure and KH SFE expressions were used.

4 Results

4.1 Unparameterized calculations

COSMO$_{ADF}$, 3DRISM-SCF$_{ADF}$ and standalone 3DRISM$_{RISM-MOL}$ calculations were performed for all the molecules in the set.

Non-corrected results of calculations one can find in Table 6. It is known, that 3DRISM typically overestimates SFEs [1]. We can see this for standalone calculations with MDC-m, MDC-d, and CHELPG charges. For all these three methods the minimal calculated SFE is greater than 0, and the maximal calculated SFE is about 30 kcal/mol, while experimental measurements lie in interval $-9.5\text{kcal/mol} \leq \Delta G_{exp} \leq 3.14\text{kcal/mol}$. However, results of 3DRISM-SCF$_{ADF}$ and standalone 3DRISM$_{RISM-MOL}$ with MDC-q charges calculations differ from the results of other 3DRISM methods. Among the reasonable results there are extremely small SFEs, like $-3.3 \cdot 10^6$ kcal/mol for 3DRISM-SCF$_{ADF}$ and -3634 kcal/mol for standalone 3DRISM$_{RISM-MOL}$ with MDC-q charges, which are obviously incorrect. Considering, that 3DRISM typically overestimates SFE we define as unreliable all those results of 3DRISM calculations, where SFE is smaller than the minimal experimentally measured SFE (-9.5 kcal/mol). These values are selected with the bold font in the Table 6. We see, that for the molecules with unreliable standalone 3DRISM$_{RISM-MOL}$ MDC-q

[1] In the Table 6 the short names of the methods are used. "SCF" stands for "3DRISM-SCF calculations", "MDC-m", "MDC-d", "MDC-q", "CHELPG" stand for "Standalone 3DRISM$_{RISM-MOL}$ calculations with MDC-m, MDC-d, MDC-q, CHELPG charges". We will also from time to time use these short names to refer the methods in the methods in the text.

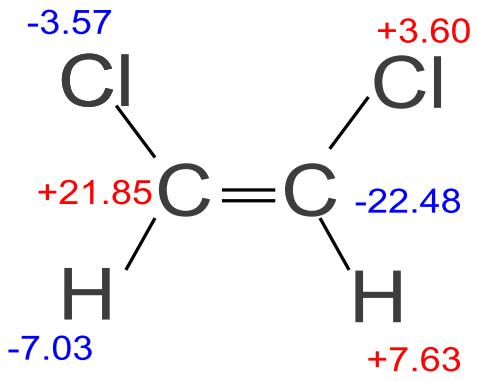

Figure 1: MDC-q partial charges of cis-1-2-dichloroethene (atomic units)

results the results of 3DRISM-SCF$_{ADF}$ calculations are also unreliable. And vice versa: for the molecules with unreliable 3DRISM-SCF$_{ADF}$ results the results of the 3DRISM$_{RISM-MOL}$/MDC-q calculations also differ much from the results of other 3DRISM methods for the same molecule. This allows us to make the following assumptions:

1. The 3DRISM-SCF$_{ADF}$ procedure uses MDC-q charges.

2. MDC-q charges are not reliable at least for some molecules in the set.

The first assumption is difficult to check without the access to the ADF source code. To check the second assumption we can check the MDC-q charges of the most problematic compounds. On the Figure 1 the charges of the cis-1-2-dichloroethene are shown. We see, that all the partial charges are greater than 3e by absolute value and some of them are greater than 20e. This is completely unphysical. Thus, we conclude, that MDC-q procedure is not suitable at least for some compounds in the set. To select the molecules with the unphysical partial charges we calculated the maximal absolute values of partial charges for each molecule in the set (see Table 7). We see, that for 23 molecules the maximal absolute value of MDC-q charges are larger than 1. Because typically partial charges of uncharged molecules are smaller than 1 by absolute value, we exclude these molecules from consideration in the current study, and investigate the rest 76 compounds.

Method	a_1	a_1^{alt}	$\lvert a_1 - a_1^{alt}\rvert$	a_2	a_2^{alt}	$\lvert a_2 - a_2^{alt}\rvert$
MDC-m (RISM-MOL)	-3.27	-3.89	0.62	-1.94	1.54	3.48
MDC-d (RISM-MOL)	-3.74	-3.39	0.349	1.82	-0.157	1.98
MDC-q (RISM-MOL)	-3.36	-3.09	0.275	3.02	0.348	2.67
CHELPG (RISM-MOL)	-3.6	-3.73	0.129	-0.274	0.484	0.758
SCF (ADF)	-3.29	-2.49	0.804	3.54	-2.18	5.72

Table 2: Coefficients in the UC model for different 3DRISM calculations and two different training sets

Method	b	b^{alt}	$\lvert b - b^{alt}\rvert$
MDC-m (RISM-MOL)	-3.57	-3.65	0.0741
MDC-d (RISM-MOL)	-3.46	-3.42	0.0387
MDC-q (RISM-MOL)	-2.89	-3.04	0.15
CHELPG (RISM-MOL)	-3.64	-3.66	0.011
SCF (ADF)	-2.74	-2.82	0.0884

Table 3: Coefficient in the UC_0 model for different 3DRISM calculations and two different training sets

4.2 Partial molar volume parameterization

The uncorrected results of 3DRISM-SCF$_{ADF}$ and standalone 3DRISM$_{RISM-MOL}$ calculations are presented in Figure **??** We see, that all 3DRISM methods considerably overestimate the SFE, and thus need to be corrected. We use the UC model to correct both: 3DRISM-SCF$_{ADF}$ and standalone 3DRISM$_{RISM-MOL}$ calculation results. We do not use the UC model for the COSMO$_{ADF}$ calculations, because it is already fitted to match the experimental data and cannot be essentially improved with the partial molar volume correction.

To use the UC model (1) we need to determine first the free coefficients a_1, a_2. To do this, we selected the training set of 20 compounds (the first 20 compounds in Table 7 with maximal absolute MDC-q charges smaller than 1). The rest of compounds with the maximal absolute MDC-q charges smaller than 1 we included to the test set (see Table 7). We performed a parameterization on the training set of compounds and found the coefficients a_1, a_2 in the formula (1). To check, whether the coefficients really reflect some physical quantities, we performed an alternative parameterization with another training set. The alternative training set includes the

Method	UC						UC_0			
	a_1		a_2		RMSD		b		RMSD	
MDC-m	-3.66	±0.44	-0.216	±2.70	3.36	±0.20	-3.7	±0.10	3.27	±0.15
MDC-d	-3.53	±0.15	0.674	±0.91	1.15	±0.07	-3.42	±0.04	1.12	±0.06
MDC-q	-3.32	±0.44	2.69	±2.56	3.34	±0.22	-2.9	±0.11	3.31	±0.19
CHELPG	-3.56	±0.14	-0.199	±0.91	1.12	±0.07	-3.59	±0.03	1.09	±0.05
SCF	-3	±0.70	2.07	±3.96	5.59	±0.40	-2.68	±0.17	5.47	±0.36

Table 4: Coefficients and RMSD errors for UC and UC_0 models and different 3DRISM methods. The values are given in kcal/mol in the format mean value ± standard deviation.

last 20 compounds in Table 7 with the maximal absolute MDC-q charges smaller than 1. In such a way we calculated the coefficients a_1^{alt} and a_2^{alt}. In Table 2 the coefficients a_1, a_2, a_1^{alt}, a_2^{alt} are presented. We see, that the coefficients essentially depend on the choice of the training set, and the coefficient a_2 changes much more than a_1. This allows us to suggest that the coefficient a_2 does not reflect any physical quantity. To check this assumption we perform the similar parameterizations with the zero a_2 coefficient. We will refer this parameterization as the UC_0-model. The following expression is used to calculate the SFE in the UC_0-model:

$$\Delta G_{UC_0} = \Delta G_{RISM} + b\rho V \tag{2}$$

where the coefficient b is calculated by the least squares method on a training set of compounds. We perform the same analysis for the UC_0 model as we have done for the UC model. In Table 3 the coefficients b for two different training sets and different 3DRISM methods are listed. We can see, that for most of the methods the difference $|b - b^{alt}|$ is smaller than 0.1, which is much less than the same differences in the UC model. This allows us to suggest that UC_0 model is much more stable and less dependent on the training set.

To reduce the probability of mistake we performed additional calculations for 1000 different training sets. Each of these training sets contained 20 randomly chosen molecules from the 76 molecules listed in Table 7 with maximal MDC-q partial charges smaller than 1. For each of these training sets we calculated coefficients of the UC and UC_0 models using the least squares method. Using these coefficients for each training set we calculated the SFEs of the rest 56 molecules (not in the training set) and compare these predictions to the experimental results. We calculated the

Method	b	RMSD
MDC-m	-3.57	3.39
MDC-d	-3.46	1.13
MDC-q	-2.89	3.27
CHELPG	-3.57	1.1
SCF	-2.74	5.5
COSMO	-	4.32

Table 5: Coefficient (b) of the UC_0 models for different charges and RMSD errors of UC_0 models with different charges and COSMO.

mean value and standard deviation of the coefficients and of the RMSD errors using the data obtained for 1000 different training sets. The results for MDC-m, MDC-d, MDC-q, CHELPG and 3DRISM-SCF_{ADF} methods are presented in Table 4. We see that these results support our suggestions: the a_2 coefficients in the UC model vary considerably and do not improve predictability of the model.

So, we conclude that UC_0 model should be used for SFE calculations. We return back to the initial training and test sets (see Table 7) and using the least squares method on the training set obtain the coefficients for the UC_0 model. The results of the 3DRISM calculations with the UC_0 model and also results of the $COSMO_{ADF}$ calculations for the test set of compounds are presented in Figure 3 and in Table 5. We see, that only two methods: standalone $3DRISM_{RISM-MOL}$ calculations with MDC-d and CHELPG in combination with UC_0 model charges give reasonable predictions of SFE (accuracy about 1.1 kcal/mol). All other methods (including 3DRISM-SCF_{ADF} and $COSMO_{ADF}$) fail to predict SFE with the accuracy better than 3 kcal/mol. We note though, that the SFE calculated with $COSMO_{ADF}$ seem to be linearly dependent on the experimental data. This means that the $COSMO_{ADF}$ results can be parameterized using the linear regression method.

5 Conclusions

We benchmarked the ADF 3DRISM-SCF and COSMO tools on the set of 99 organic compounds. Additionally we performed the standalone $3DRISM_{RISM-MOL}$ calculations for the same compounds using four different partial charges: MDC-m, MDC-d, MDC-q, CHELPG. We have shown, that MDC-q charges are unreliable for almost quoter of compounds. We have shown, that abnormal results of the standalone $3DRISM_{RISM-MOL}$ calculations with MDC-q charges are correlated with the abnor-

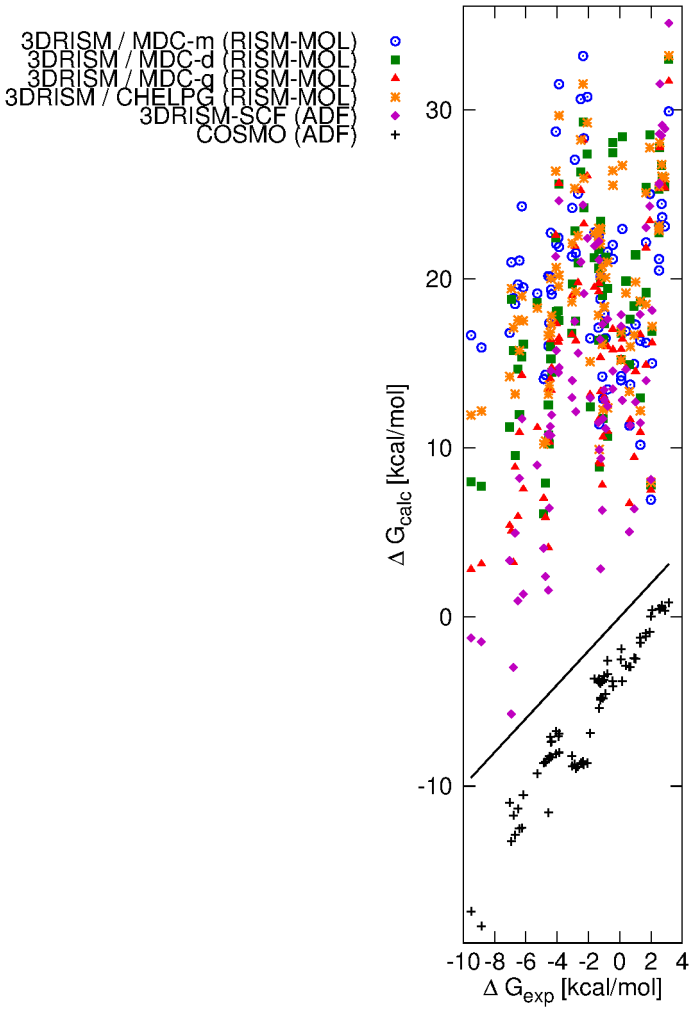

Figure 2: Solvation free energies calculated with the 3DRISM-SCF$_{ADF}$, standalone 3DRISM$_{RISM-MOL}$ and COSMO$_{ADF}$ methods. "MDC-m", "MDC-d", "MDC-q", "CHELPG" labels correspond to the standalone 3DRISM$_{RISM-MOL}$ calculations with different partial charges. "SCF" label corresponds to the 3DRISM-SCF$_{ADF}$ calculations. "COSMO" label corresponds to the COSMO$_{ADF}$ calculations.

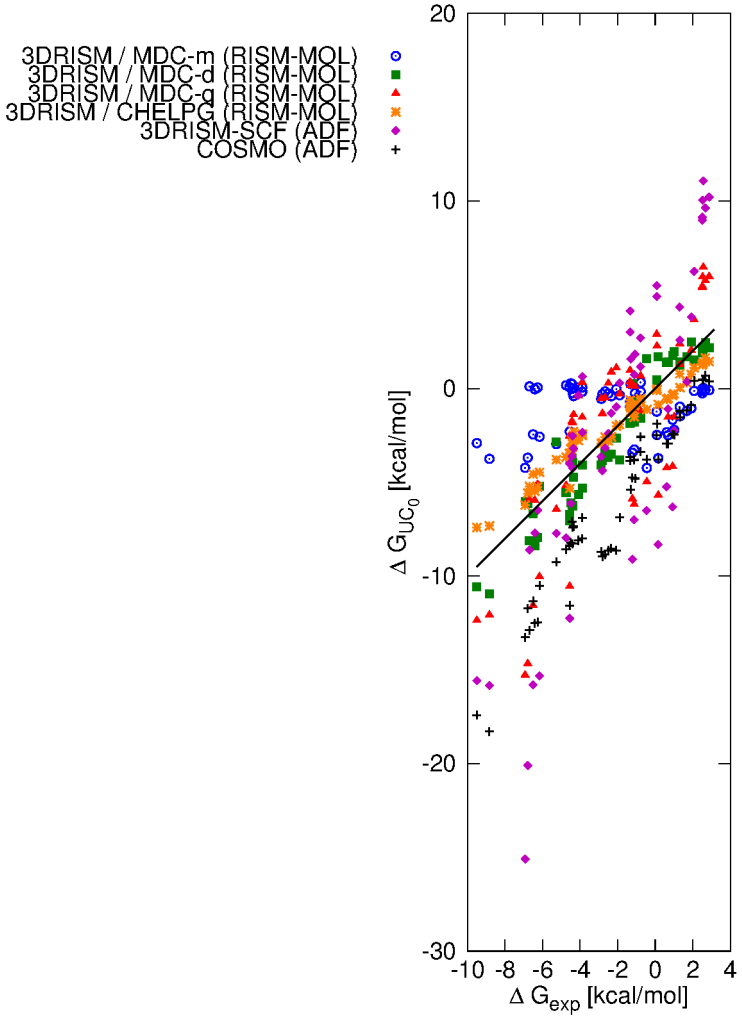

Figure 3: Solvation free energies calculated with the 3DRISM-SCF$_{ADF}$ and standalone 3DRISM$_{RISM-MOL}$ methods and corrected with the UC$_0$. Uncorrected results of the COSMO$_{ADF}$ solvation free energy calculations. "MDC-m", "MDC-d", "MDC-q", "CHELPG" labels correspond to the standalone 3DRISM$_{RISM-MOL}$ calculations with different partial charges. "SCF" label corresponds to the 3DRISM-SCF$_{ADF}$ calculations. "COSMO" label corresponds to the COSMO$_{ADF}$ calculations.

mal results of the 3DRISM-SCF$_{ADF}$ calculations. We suggested, that MDC-q charges are used in the 3DRISM-SCF$_{ADF}$ calculations. For the molecules with reasonable MDC-q charges we performed parameterization using the UC model. We have shown that the coefficient a_2 in the UC model strongly depends on the choice of the training set and does not improve the accuracy of the SFE calculations. We proposed the UC$_0$ model with only one PMV coefficient. We have shown that the standalone 3DRISM$_{RISM-MOL}$ calculations with UC$_0$ model for MDC-d and CHELPG charges give the accuracy of SFE calculations of about 1.1 kcal/mol, while other methods (including 3DRISM-SCF$_{ADF}$ and COSMO$_{ADF}$) fail to predict SFEs with accuracy better than 3 kcal/mol.

6 Discussion

As it was discussed above, we have found the correlation between the failures of standalone 3DRISM$_{RISM-MOL}$ calculations with MDC-q charges and failures of 3DRISM-SCF$_{ADF}$ method. This allowed us to suggest that MDC-q charges are used in the 3DRISM-SCF$_{ADF}$ procedure. If our suggestion is correct, then correction of the 3DRISM-SCF$_{ADF}$ code should not be too complicated: it is enough to use other charges in 3DRISM-SCF$_{ADF}$ procedure. Good candidates are MDC-d or CHELPG charges. We should note though, that partial charges in 3DRISM calculations are used only for calculation of the electrostatic potential at the grid points around the molecule. In QM calculations electrostatic potential can be calculated directly, without calculation of the partial charges, and this potential can be directly inserted into the 3DRISM calculations. This will allow one to avoid problems with partial charges calculation and to achieve better accuracy of calculations.

References

[1] Volodymyr .P Sergiievskyi and Maxim V. Fedorov. 3drism multigrid algorithm for fast solvation free energy calculations. *Journal of Chemical Theory and Computation*, 8(6):2062–2070, 2012.

[2] Rism-mol3d: Multi-grid 3drism equations solver, http://compchemmpi.wikispaces.com/RISM-MOL-3D.

[3] S. Palmer, A. I. Frolov, E. L. Ratkova, and M. V. Fedorov. Towards a universal method to calculate hydration free energies: a 3d reference interaction site model with partial molar volume correction. *Journal of Physics: Condensed Matter*, 22(49):492101, 2010.

[4] Marcel Swart, Piet Th. van Duijnen, and Jaap G. Snijders. A charge analysis derived from an atomic multipole expansion. *Journal of Computational Chemistry*, 22(1):79–88, 2001.

[5] D. S. Palmer, V. P. Sergiievskyi, F. Jensen, and M. V. Fedorov. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *Journal of Chemical Physics*, 133(4):044104, July 2010.

[6] V. P. Sergiievskyi. Model for calculating the free energy of hydration of bioactive compounds based on integral equations of the theory of liquids. *Russian Journal of Physical Chemistry B*, 5(2):326–331, 2011.

[7] E. L. Ratkova, G. N. Chuev, V. P. Sergiievskyi, and M. V. Fedorov. An accurate prediction of hydration free energies by combination of molecular integral equations theory with structural descriptors. *Journal of Physical Chemistry B*, 114(37):12068–12079, 2010.

[8] Orca user manual, www.mpibac.mpg.de/bac/logins/neese/downloads/OrcaManual_2_9.pdf.

A Supporting information

Table 6: Solvation Free Energies calculated with different methods (raw data in kcal/mol).

No	Compound	COSMO	SCF	MDC-m	MDC-d	MDC-q	CHELPG	Exp
1	1-1-dichloroethane	-4.65	**-10.18**	13.99	11.36	8.008	13.1	-0.846
2	1-1-2-trichloroethane	-9.58	7.204	16.46	12.13	10.53	13.75	-1.991
3	1-2-dichloropropane	-7	12.48	17.15	12.53	12.72	14.94	-1.269
4	1-2-3-5-tetrachlorobenzene	-3.65	21.95	22.75	21.27	19.5	22.69	-1.623
5	1-3-dichlorobenzene	-3.48	12.84	17.95	17.36	12.41	18.35	-0.982
6	1-4-dichlorobenzene	-3.7	13.43	17.94	16.38	13.01	18.35	-1.009
7	2-ethyltoluene	-4.29	7.655	19.71	24.19	11.81	22.71	-1.037
8	2-methylpentan-2-ol	-7.06	14.47	22.45	18.11	16.26	19.58	-3.927
9	2-methylstyrene	-4.89	2.852	18.81	23.01	9.059	22.06	-1.24
10	2-3-dimethylbuta-1-3-diene	-2.88	14.64	16.92	19.88	14.73	19.17	0.394
11	2-4-dimethylphenol	-10.2	-3.088	20.51	17.42	3.606	17.96	-6.013
12	3-methylhexane	0.58	29.1	23.66	25.98	25.86	26.08	2.713
13	4-chlorophenol	-10.98	3.337	16.81	11.25	5.409	14.23	-7.036
14	4-methylpentan-2-one	-8.22	12.99	21.35	16.79	16.7	18.65	-3.054
15	cis-1-2-dichloroethene	-3.78	**-1.34e+09**	13.15	11.47	**-3634**	11.9	-1.174
16	heptan-2-one	-8.81	14	24.21	19.72	19.02	22.11	-3.04
17	hexan-3-ol	-6.76	15.77	22.08	18.08	17.36	20.68	-4.063
18	methane	0.05	8.138	6.946	7.821	7.515	7.973	1.991
19	n-nonane	0.86	35.14	29.92	32.99	31.69	33.21	3.136
20	o-xylene	-4.55	11.15	16.5	21.3	12.83	20.07	-0.901
21	octanal	-8.72	19.14	28.33	24.22	23.25	25.97	-2.292
22	pentachloroethane	-4.43	**-6696**	19.43	18.47	**-26.55**	18.85	-1.391
23	propan-1-ol	-8.63	4.074	14.09	6.118	7.026	10.24	-4.854
24	tert-butylbenzene	-4.09	13.49	21.19	27.47	15.81	25.55	-0.437
25	trans-hept-2-ene	-0.97	23.05	22.17	25.41	21.84	25.1	1.678
26	1-1-dichloroethene	-2.15	**-9.962e+04**	12.85	12.23	-0.8453	12.79	0.246
27	1-1-1-trichloroethane	-2.89	12.6	15.49	14.6	11.34	15.61	-0.191
28	1-1-1-2-tetrachloroethane	-4.89	16.47	17.91	15.92	15.51	17.08	-1.281
29	1-1-2-2-tetrachloroethane	-5.38	**-1.445e+06**	18.36	14.54	**-355.2**	16.88	-2.469
30	1-2-dichlorobenzene	-3.78	11.5	17.12	16.38	11.64	17.86	-1.365
31	1-2-dichloroethane	-5.16	**-18.07**	14.02	10.32	-0.4872	12.97	-1.785

Continued on the next page

No	Compound	COSMO	SCF	MDC-m	MDC-d	MDC-q	CHELPG	Exp
32	1-2-3-trichlorobenzene	-3.92	16.44	20.14	19.79	15.35	20.21	-1.24
33	1-2-3-trimethylbenzene	-4.76	9.386	20	23.41	13.35	22.93	-1.214
34	1-2-3-4-tetrachlorobenzene	-3.85	21.13	22.55	21.76	19.25	22.55	-1.336
35	1-2-4-trichlorobenzene	-3.79	17.32	21.15	19.01	16.08	20.74	-1.119
36	1-2-4-trimethylbenzene	-4.64	0.6643	21.35	23.99	8.079	23.14	-0.858
37	1-2-4-5-tetrachlorobenzene	-3.66	22.22	22.79	20.64	19.73	23	-1.336
38	1-3-dichloropropane	-6.87	12.97	16.49	12.42	13.14	15.11	-1.895
39	1-3-5-trichlorobenzene	-3.37	17.62	21.61	19.46	15.98	21	-0.777
40	1-3-5-trimethylbenzene	-4.48	**-10.84**	22.11	24.45	1.428	23.83	-0.901
41	2-butoxyethanol	-12.47	11.73	24.3	15.41	14.3	18.98	-6.26
42	2-chlorophenol	-11.57	1.595	16.04	10.75	4.093	13.18	-4.555
43	2-ethoxyethanol	-12.89	4.974	18.52	9.557	8.857	13.19	-6.697
44	2-methylbut-2-ene	-1.21	17.9	16.35	18.65	16.7	18.67	1.31
45	2-methylbuta-1-3-diene	-2.95	11.34	13.75	17.62	11.67	16	0.681
46	2-methylbutan-2-ol	-7.09	11.28	20.11	15.26	14.08	16.83	-4.431
47	2-methylpentan-3-ol	-6.89	14.76	21.9	17.55	16.49	20.19	-3.886
48	2-methylpentane	0.49	25.6	21.18	23.06	22.91	23.16	2.51
49	2-methylpropan-1-ol	-8.23	6.456	17.39	10.24	10.34	13.71	-4.5
50	2-phenylethanol	-11.73	-2.961	18.87	15.76	3.221	17.12	-6.793
51	2-propoxyethanol	-12.51	8.208	21.09	11.99	10.91	15.76	-6.41
52	2-2-dimethylpentane	0.39	28.89	23.12	25.47	25.39	25.98	2.878
53	2-3-dimethylpentane	0.46	28.59	23.18	25.31	25.36	25.47	2.524
54	2-3-dimethylphenol	-10.52	1.362	19.51	16.15	7.567	17.53	-6.164
55	2-3-4-trimethylpentane	0.46	31.56	25.51	27.76	27.77	28.07	2.565
56	2-5-dimethylphenol	-10.71	-8.703	20.52	17.69	0.1089	18.36	-5.918
57	2-6-dimethylphenol	-9.25	8.989	19.14	18.59	11.2	18.28	-5.265
58	3-hydroxybenzaldehyde	-17.42	-1.222	16.67	8.007	2.816	11.96	-9.505
59	3-methylpentane	0.46	25.68	20.51	22.73	22.93	22.88	2.51
60	3-phenylpropanol	-13.27	-5.723	20.99	18.81	5.075	19.43	-6.929
61	3-4-dimethylphenol	-11.34	0.9651	19.68	14.66	5.935	17.58	-6.506
62	4-ethyltoluene	-4.21	5.171	20.9	24.97	10.23	23.91	-0.954
63	4-hydroxybenzaldehyde	-18.28	-1.466	15.95	7.747	3.131	12.19	-8.836
64	4-methoxyacetophenone	-13.57	9.577	20.24	15.06	13.36	19.46	-4.405
65	benzyl-alcohol	-12.89	**-2.3e+04**	15.39	12.08	-7.77	14.49	-6.628
66	buta-1-3-diene	-2.94	5.038	11.33	14.92	6.713	13.34	0.614
67	chlorobenzene	-3.82	6.326	14.23	16.46	7.804	16.13	-1.119

Continued on the next page

No	Compound	COSMO	SCF	MDC-m	MDC-d	MDC-q	CHELPG	Exp
68	decan-2-one	-8.54	24.38	33.2	29.28	28.23	31.54	-2.345
69	dichloromethane	-5.39	9.907	11.41	8.89	9.177	9.919	-1.31
70	ethane	0.25	**-2.597e+08**	9.417	10.88	**-21.16**	11.07	1.828
71	heptanal	-8.86	15.62	25.06	20.96	19.77	22.57	-2.672
72	hexa-1-5-diene	-2.47	12.74	17.31	21.43	14.52	19.82	1.009
73	hexan-1-ol	-8.28	14.58	22.74	16.01	16.4	20.02	-4.405
74	hexanal	-8.95	12.15	21.54	17.5	16.33	19.24	-2.808
75	isobutylbenzene	-3.79	12.83	22.96	28.41	16.43	26.72	0.163
76	m-xylene	-4.39	-8.336	17.5	21.57	-1.172	20.4	-0.832
77	methanol	-9.65	**-14.74**	8.262	0.2966	-3.257	4.134	-5.1
78	n-butane	0.41	18.15	15.02	16.91	16.21	17.2	2.072
79	n-heptane	0.69	28.48	24.45	26.71	25.83	26.77	2.672
80	nonan-1-ol	-8.01	24.63	31.53	25.62	25.66	29.66	-3.886
81	nonan-2-one	-8.63	21.01	30.65	26.34	25.23	28.23	-2.495
82	nonanal	-8.64	22.41	30.78	27.4	26.08	29.25	-2.072
83	oct-1-ene	-0.89	24.31	25.03	28.52	23.43	27.77	1.924
84	octan-1-ol	-8.1	21.33	28.72	22.42	22.58	26.38	-4.092
85	octan-2-one	-8.71	17.49	27.05	22.85	21.95	25.36	-2.878
86	p-xylene	-4.33	**-28.13**	18.07	21.73	**-11.98**	20.36	-0.805
87	pent-1-ene	-1.16	13.99	16.24	19.2	14.91	18.49	1.678
88	penta-1-4-diene	-2.43	6.4	14.97	18.4	9.447	16.69	0.927
89	pentan-1-ol	-8.39	10.81	20.17	12.55	13.44	16.65	-4.57
90	pentan-2-ol	-7.4	10.75	19.37	15.29	13.42	17.32	-4.391
91	pentan-3-ol	-7.36	11.95	19.1	14.48	14.73	17.83	-4.35
92	propan-2-ol	-8.58	2.394	14.34	7.916	5.891	10.41	-4.747
93	propene	-1.53	11.5	10.19	12.96	10.93	12.2	1.322
94	sec-butylbenzene	-3.79	14.57	22.01	28.08	17.03	26.4	-0.449
95	tetrachloroethene	-1.88	17.89	14.26	16.81	15.83	16.88	0.096
96	tetrachloromethane	-2.51	17.2	14.01	15.25	15.25	15.2	0.081
97	trans-1-2-dichloroethene	-2.58	12.48	13.47	10.7	10.98	12.38	-0.777
98	trichloroethene	-2.48	**-356.1**	14.65	13.92	7.694	14.61	-0.437
99	trichloromethane	-4.8	11.74	12.92	11.76	10.64	12.27	-1.078

Table 7: Maximal absolute partial charges calculated with the different methods. Training, test, and alternative training sets.

No	Compound	MDC-m	MDC-d	MDC-q	CHELPG	Training	Test	Alt. Training
1	1-1-dichloroethane	0.8642	0.3397	**1.512**	0.2366			
2	1-1-2-trichloroethane	0.4268	0.2038	**1.104**	0.2252			
3	1-2-dichloropropane	0.832	0.2645	**1.507**	0.3419			
4	1-2-3-5-tetrachlorobenzene	0.3758	0.2374	0.2751	0.121	*		
5	1-3-dichlorobenzene	0.3611	0.1945	0.2763	0.1766	*		
6	1-4-dichlorobenzene	0.2796	0.1681	0.2496	0.1173	*		
7	2-ethyltoluene	0.8797	0.312	**1.155**	0.4467			
8	2-methylpentan-2-ol	0.9246	0.6467	0.7839	0.6878	*		
9	2-methylstyrene	0.8757	0.3077	0.6684	0.3999	*		
10	2-3-dimethylbuta-1-3-diene	0.8735	0.3168	0.5364	0.4966	*		
11	2-4-dimethylphenol	0.891	0.5675	**2.917**	0.5414			
12	3-methylhexane	0.8624	0.2838	0.4242	0.318	*		
13	4-chlorophenol	0.288	0.6086	0.6311	0.5615	*		
14	4-methylpentan-2-one	0.9735	0.4699	0.5717	0.6994	*		
15	cis-1-2-dichloroethene	0.07994	0.1371	**22.48**	0.162			
16	heptan-2-one	0.9698	0.4664	0.5557	0.6821	*		
17	hexan-3-ol	0.8397	0.6912	0.7183	0.603	*		
18	methane	**1.153**	0.3691	0.7277	0.4567	*		
19	n-nonane	0.8328	0.2292	0.5554	0.2544	*		
20	o-xylene	0.887	0.3195	0.8612	0.3888	*		
21	octanal	0.8321	0.3697	0.5345	0.5107	*		
22	pentachloroethane	0.4939	0.266	**2.749**	0.3902			
23	propan-1-ol	0.7963	0.7967	0.9949	0.6441	*		
24	tert-butylbenzene	0.9087	0.34	0.4582	0.6553	*		
25	trans-hept-2-ene	0.8699	0.2788	0.5013	0.2662	*		
26	1-1-dichloroethene	0.5971	0.2184	**1.892**	0.3271			
27	1-1-1-trichloroethane	0.9018	0.4246	**2.211**	0.4157			
28	1-1-1-2-tetrachloroethane	0.5447	0.2085	**1.419**	0.3694			
29	1-1-2-2-tetrachloroethane	0.02686	0.3184	**10.46**	0.2605			
30	1-2-dichlorobenzene	0.3043	0.2714	0.6947	0.1255	*		
31	1-2-dichloroethane	0.3976	0.2641	**1.657**	0.1538			
32	1-2-3-trichlorobenzene	0.3018	0.1067	0.2544	0.1545	*		

Continued on the next page

No	Compound	MDC-m	MDC-d	MDC-q	CHELPG	Training	Test	Alt. Training
33	1-2-3-trimethylbenzene	0.8859	0.323	0.42	0.3537		*	
34	1-2-3-4-tetrachlorobenzene	0.2836	0.09631	0.2533	0.1295		*	
35	1-2-4-trichlorobenzene	0.3735	0.198	0.2747	0.1403		*	
36	1-2-4-trimethylbenzene	0.8913	0.3201	**2.686**	0.437			
37	1-2-4-5-tetrachlorobenzene	0.3849	0.1605	0.2585	0.12		*	
38	1-3-dichloropropane	0.4976	0.2555	0.4193	0.214		*	
39	1-3-5-trichlorobenzene	0.3666	0.2181	0.2921	0.107		*	
40	1-3-5-trimethylbenzene	0.8914	0.3139	**2.445**	0.5638			
41	2-butoxyethanol	0.8337	0.8004	0.8093	0.6687		*	
42	2-chlorophenol	0.3162	0.5722	0.5888	0.5649		*	
43	2-ethoxyethanol	0.8227	0.8011	0.814	0.6553		*	
44	2-methylbut-2-ene	0.9036	0.3433	0.5079	0.3511		*	
45	2-methylbuta-1-3-diene	0.8667	0.3133	0.7803	0.5665		*	
46	2-methylbutan-2-ol	0.9233	0.6462	0.7749	0.6915		*	
47	2-methylpentan-3-ol	0.8547	0.7044	0.759	0.6576		*	
48	2-methylpentane	0.8662	0.2852	0.4282	0.3976		*	
49	2-methylpropan-1-ol	0.8554	0.7882	0.7969	0.6112		*	
50	2-phenylethanol	0.5448	0.7897	0.8048	0.6643		*	
51	2-propoxyethanol	0.8121	0.7999	0.81	0.6676		*	
52	2-2-dimethylpentane	0.9004	0.3334	0.3697	0.5945		*	
53	2-3-dimethylpentane	0.8653	0.2823	0.3217	0.3445		*	
54	2-3-dimethylphenol	0.8718	0.5724	0.6617	0.5487		*	
55	2-3-4-trimethylpentane	0.8587	0.278	0.341	0.4732		*	
56	2-5-dimethylphenol	0.9019	0.5847	**3.298**	0.539			
57	2-6-dimethylphenol	0.9026	0.5459	0.8827	0.5176		*	
58	3-hydroxybenzaldehyde	0.3034	0.5989	0.6304	0.552		*	
59	3-methylpentane	0.8638	0.2859	0.2603	0.2943		*	
60	3-phenylpropanol	0.5198	0.7705	0.9273	0.6705		*	
61	3-4-dimethylphenol	0.8869	0.655	0.6399	0.5568		*	
62	4-ethyltoluene	0.8902	0.3106	**1.927**	0.3785			
63	4-hydroxybenzaldehyde	0.2865	0.5903	0.6223	0.5441		*	
64	4-methoxyacetophenone	0.9605	0.455	**1.192**	0.7055			
65	benzyl-alcohol	0.5515	0.7362	**2.11**	0.6094			
66	buta-1-3-diene	0.5095	0.1021	0.5131	0.4258		*	
67	chlorobenzene	0.2854	0.1676	0.2621	0.1402		*	
68	decan-2-one	0.9697	0.4663	0.5238	0.6946		*	

Continued on the next page

No	Compound	MDC-m	MDC-d	MDC-q	CHELPG	Training	Test	Alt. Training
69	dichloromethane	0.3014	0.1488	0.5988	0.1778		*	
70	ethane	0.8141	0.2033	**2.446**	0.00275			
71	heptanal	0.8325	0.3701	0.5345	0.4866		*	
72	hexa-1-5-diene	0.5316	0.2073	0.5587	0.4455		*	
73	hexan-1-ol	0.8321	0.7926	0.9495	0.6486		*	
74	hexanal	0.8319	0.3701	0.5394	0.4849		*	
75	isobutylbenzene	0.8647	0.2833	0.334	0.5694		*	*
76	m-xylene	0.89	0.3126	**2.564**	0.4907			
77	methanol	0.8904	0.8307	**1.842**	0.6202			
78	n-butane	0.828	0.224	0.5384	0.3082		*	*
79	n-heptane	0.832	0.2292	0.5331	0.2742		*	*
80	nonan-1-ol	0.8322	0.7926	0.9523	0.652		*	*
81	nonan-2-one	0.9697	0.4661	0.5283	0.7133		*	*
82	nonanal	0.8326	0.3699	0.5342	0.5007		*	*
83	oct-1-ene	0.8328	0.227	0.553	0.4415		*	*
84	octan-1-ol	0.8322	0.7926	0.9565	0.6526		*	*
85	octan-2-one	0.9698	0.4665	0.5174	0.6895		*	*
86	p-xylene	0.891	0.3116	**3.995**	0.3479			
87	pent-1-ene	0.8253	0.2151	0.5612	0.4262		*	*
88	penta-1-4-diene	0.5463	0.1727	0.6281	0.4677		*	*
89	pentan-1-ol	0.8311	0.7928	0.9414	0.6572		*	*
90	pentan-2-ol	0.8712	0.6941	0.7987	0.6702		*	*
91	pentan-3-ol	0.8273	0.7239	0.7341	0.6249		*	*
92	propan-2-ol	0.8806	0.7047	0.9989	0.6732		*	*
93	propene	0.8695	0.2873	0.6019	0.5374		*	*
94	sec-butylbenzene	0.8679	0.282	0.4178	0.3158		*	*
95	tetrachloroethene	0.2662	0.04109	0.1688	0.002836		*	*
96	tetrachloromethane	0.7082	0.2255	0.2285	0.2702		*	*
97	trans-1-2-dichloroethene	0.08149	0.2453	0.2904	0.1806		*	*
98	trichloroethene	0.3314	0.1891	**1.434**	0.168			
99	trichloromethane	0.185	0.2613	0.5677	0.2227		*	*
	Average	0.7035	0.401	**1.218**	0.435			
	Max	**1.153**	0.8307	**22.48**	0.7133			